Patrick Frank 8 October 2017 Earth and Space Science Manuscript 2017EA000308 Response to Round 2 Reviewer #1 (round 1 reviewer #3):

This reviewer:

- Irrelevantly fixated on the Stefan-Boltzmann equation (items 1.2.1, 1.2.2, and 1.4)
- Consistently confused statistical uncertainty with physical error (items 1.3.3, 2.7, 4.1, 4.2.2, and 5.1)
- Invariably confused the statistical uncertainty in temperature with a physically real temperature (items 1.3.1, 2.4, and 4.2.2)
- Mistook calibration uncertainty to be the slope of a trend (items 2.3.1, 2.3.2, 2.4-2.6)
- Ignored the dimensional analysis of LWCF calibration uncertainty (ms Sect. 2.4.1, SI Sect. 6.2), and evidently does not understand that statistical averages are always per unit averaged (items 2.2 and 2.6)
- Apparently does not understand the difference between a magnitude average and a statistical average (items 2.1, 2.2, 2.4, and 2.6)
- Is refuted by his own cited authority on cloud forcing (item 3.1.1)
- Completely neglected the extensive discussion of uncertainty (item 4.1)
- Apparently does not understand calibration at all (items 1.3.1, 1.3.2, 1.3.3, 2.3.1, 2.3.2, 2.4, 2.5, 2.6, 2.7, 3.1.1, 3.1.2, 4.2.2, and 5.1)

In summary, these reviewer mistakes remove any critical merit from this review.

Detailed responses are below. The reviewer is quoted in italics, followed by the indented response.

- 1.1 I insist my opinion that this MS is scientifically wrong and cannot be published. Here is my response to the author's reply.
 - 1.1 Insistence has no force in a scientific debate. As shown below, this reviewer's criticism is thoroughly misguided.
- 1.2 From opinion on the S-B relationship, I can interpret that the author is not really understanding radiation. No matter climate physics or climate model, the S-B relation is intrinsic in the radiation budget.
 - 1.2.1 In the first round, this reviewer pointed out that the S-B relation includes temperature to the fourth power. The author did not dispute this. He merely pointed out that manuscript eqn. 6 successfully emulated the behavior of climate models, and that the S-B relation had no bearing on that emulation.

The reviewer's fixation on the S-B relation is thus critically irrelevant.

The author's point remains true, and does not indicate any lack of understanding.

- 1.2.2 If anything, the reviewer's recapitulation of this point merely shows that the reviewer does not understand that the S-B relation does not determine the air temperature of the terrestrial climate.
- 1.2.3 The reviewer also apparently does not understand the significance of eqn. 6, or the significance of its invariably successful emulation of climate model air temperature projections. This, despite the author having explicitly directed the reviewer's attention to this demonstration in round 1 response item 1.

For the benefit of this reviewer, the author further points out that whereas the S-B relationship is indeed intrinsic to the radiation budget, the S-B relation is not a complete theory of climate. Therefore, while it has bearing on air temperature, the S-B relation is not known to determine terrestrial air temperature.

- 1.3 The linear relation of Pyle et al., 2016 is only valid when the change of temperature is very small to its own value, say ~1% meaning 3 K change for a 300 K climatology. For the error range in this paper, if 30 K error appears in the change, it is already 10% of the climatology, for which lamda is never a constant.
 - 1.3.1 The reviewer is here confounding a propagated uncertainty in temperature, e.g., a ±15 K statistic, with a physically real spread of physically real temperatures. A temperature uncertainty statistic is not a physical temperature.

This absolutely fundamental mistake completely removes any critical force from reviewer comment 1.3.

1.3.2 Further, one merely need inspect manuscript Figures 2, 3, 4, 8, and 9 to notice that all the emulation temperature values are in the "climatology" range required by the reviewer (~3 K).

Therefore the manuscript use of eqn. 6 exactly meets the reviewer's required limit, despite that the reviewer evidently does not know this.

Once again, the reviewer's argument here represents the plus/minus uncertainty in temperature as though it were the calculated emulation temperature. It is not.

1.3.3 Finally, an uncertainty statistic is not a physical error. The reviewer's comments throughout indicate a complete lack of understanding concerning this critically central distinction.

One notes that Box 1.3 in [*Pyle et al.*, 2016] does not limit the efficacy of $\Delta T_s = \lambda \Delta F$ to 1% of climatology and that the reviewer provided no citation in support of that statement.

- 1.4 For Plank feedback, the strongest negative climate feedback, the calculation should be based on power 4 of temperature, and lamda is propositional to power 3 of temperature.
 - 1.4 This fixation on S-B and the Planck feedback stems from the reviewer's fatal misconstrual of the temperature uncertainty statistic as a physically real temperature.

The reviewer should also notice that manuscript eqn. 6 does not include temperature as an intensive quantity. That is, temperature nowhere enters the right side of eqn. 6.

The reviewer's focus on the equation in [Pyle et al., 2016] is thus irrelevant at best.

- 2. Again, the error in climatology cannot be directly counted for change. It should be as the similar percentage in magnitude. The dimension of +-4 W m-2 year-1 is not right, but should be +-4 W m-2, since it is calculated from 20-yr means but not 20-yr trend.
 - 2.1 The reviewer has completely ignored the new dimensional analysis provided in revised Section 2.4.1 and the full derivation in revised SI section 6.2. These fully demonstrate the per year denominator in the LW cloud forcing uncertainty statistic.
 - 2.2 The reviewer is factually incorrect. Consider: (20-year sum of annual uncertainties)/20 years = uncertainty per year. This is not difficult.

Nevertheless, the $\pm 4 \text{ Wm}^{-2}\text{year}^{-1}$ is the annual mean uncertainty derived from individually simulated 20-year hindcast trends in cloud cover.

Therefore, the uncertainty was indeed derived from a trend.

2.3.1 The ±4 Wm⁻²year⁻¹ is a calibration uncertainty, derived from comparison of a test result against a known standard. Calibration uncertainties are not restricted to observational trends.

It appears the reviewer does not understand the difference between a calibration uncertainty and a slope.

2.3.2 Each model annual hindcast simulation is a test result. Each past annual observed cloud cover is a known standard. There is no need these should be restricted to a sequential series of years. I.e., they only need to be pair-wise comparable.

The test-result (simulation) vs. observation pairs can be aggregated into a calibration uncertainty statistic; they need not demarcate a trend. If the calibration pairs are time-wise annual, the resulting calibration uncertainty denominator will be of dimension year⁻¹.

2.4 In general, the reviewer is conflating a statistical average with a measurement average. They are not the same at all. Nor do they have the same dimensions.

For example, the one-minute average of a single temperature measured at a rate of 2 Hz yields a temperature in units of Celsius.

The average of several measurements of the same air temperature, measured using an array of physically distinct, e.g., PRT sensors, is Celsius per sensor.

The temperature measurement calibration uncertainty, established for each PRT sensor by comparison against a known high-accuracy calibration sensor, is ±C per measurement.

Typically, each test PRT sensor will produce its own error distribution, yielding an empirical ±standard deviation of measurement uncertainty. In field experiments, this uncertainty distribution is typically non-normal [*Hubbard and Lin*, 2002].

That \pm C per measurement must be appended to every subsequent measurement by each PRT, and must be propagated through any series of calculations involving a measurement from these PRTs.

The calibration uncertainty attached to climate model simulations follows the same methodological process.

It is incredible that these distinctions must be explained to a PhD-level reviewer.

2.5 The reviewer seems to like web-based explanations. Here is a nice lesson on how to calculate a measurement average as distinct from a statistical average, appropriately using the number of students in a kindergarten class: http://study.com/academy/lesson/arithmetic-mean-definition-formula-example.html

An extract: The measured class sizes of each of these kindergartens are 26, 20, 25, 18, 20 and 23 [students]. Calculate the average number of students per class: (26+20+25+18+20+23)/6 classes = 132 students/6 classes = 22 students/class.

The statistical average of kindergarten enrollment is 22 students per class. What sense does this statistical average make without the identifying denominator, "per class"?

And yet, the reviewer insists the "per class" should not appear.

2.6. The author lacks basic knowledge on averaging.

2.6. One can only conclude, rather, that the reviewer lacks the basic knowledge to distinguish a magnitude average from a statistical average, and does not know the difference between a calibration uncertainty statistic and a slope.

A statistical average of like physical magnitudes has no meaning without the denominator indicating the aggregated dimension over which the magnitudes have been averaged. For example, the annual average of twenty years of simulated LWCF:

$$\frac{\sum_{i=1}^{20} (Wm^{-2})_i}{20 \text{ years}} = Wm^{-2}/\text{year}$$

R1.1

When an individual model mean simulation error is aggregated across twenty years, as in the [*Lauer and Hamilton*, 2013] calibration experiment, the standard way to calculate the calibration uncertainty for the individual model is the root-mean-square. Thus for the annual error:

$$\pm \sigma_{\text{model}} = \sqrt{\frac{\sum_{i=1}^{20} (\Delta W m^{-2})_i^2}{20 \text{ years}}} = \pm W m^{-2} / \text{ year}$$
R1.2

where ΔWm^{-2} is the model error, defined as the annual (simulation minus observed). When the annual root-mean-square calibration uncertainties of, e.g., a set of twenty-seven CMIP5 models, are combined then the calculated model mean calibration uncertainty is:

$$\pm \sigma_{LWCF} = \sqrt{\frac{\sum_{i=1}^{27} (Wm^{-2} / year)^2}{27}} = \pm Wm^{-2} / year$$
 R1.3

This calibration uncertainty applies to any simulation made using any CMIP5 model, and is to be propagated through every air temperature projection.

The derivation of these results was presented in Section 2.4.1 in the revised manuscript. The complete derivation in SI Section 6.2.was referenced in lines 525-526 of the revised manuscript. The reviewer has ignored them.

This is hardly rocket-science.

00

- 2.7 Thus, the error in climatological LWCF is not the error in its trend, which should be at similar percentage of calculated LWCF trend as the percentage of the error in climatological LWCF. The wrong conclusion of the author comes form that he use +-4 W m-2 as error in trends, because he artificially gives it a year-1 in unit.
 - 2.7 It should be obvious at this point that the reviewer item 2.7 is quite wrong. The LWCF error statistic is a calibration uncertainty, determined across many models and many test years.

It is not a physical error in climatological LWCF. It is the systematic uncertainty in simulated global LWCF, per model per year.

The dimensions of this statistic were thoroughly and transparently derived in manuscript Section 2.4.1 and SI Section 6.2. The only failure is in the reviewer's neglect of these demonstrations.

3.1 The source of this fault is that the author is always confusing about change and climatology. He does not know the definition of cloud forcing, as in the following link: <u>https://en.wikipedia.org/wiki/Cloud_forcing</u>

3.1.1 The reviewer's own authority refutes review item 3.1. Figure R1.1 below from the reviewer's Wiki page depicts long wave cloud forcing.



Figure R1.1, original legend: This image depicts the effects of clouds absorbing longwave rays from the Earth, which are then reemitted back to the surface. This tends to result in overall warming of the Earth.

Figure R1.1 shows down-welling long wave radiation, representing longwave cloud forcing. This radiation is directly part of the atmospheric thermal energy flux bath.

[*Lauer and Hamilton*, 2013] shows that CMIP5 climate model simulations of this identical thermal energy flux are uncertain to the extent of $\pm 4 \text{ Wm}^{-2}\text{year}^{-1}$.

The radiative forcing of CO_2 emissions is also part of the atmospheric thermal energy flux bath. CMIP5 simulations of this thermal flux bath are not accurate to better than an average ±4 Wm⁻²year⁻¹ [*Lauer and Hamilton*, 2013]. This ±4 Wm⁻²year⁻¹ average simulation uncertainty is a lower limit of resolution of atmospheric thermal flux for CMIP5 climate models.

Therefore, the simulated change in atmospheric thermal flux imposed by the radiative forcing of CO_2 emissions cannot be resolved to better than ±4 Wm⁻²year⁻¹.

This reasoning is entirely standard, and should not tax the understanding of any qualified physical scientist.

3.1.2 The radiative forcing of CO_2 emissions averages about 0.035 Wm⁻²year⁻¹ (since 1979), which is about 114× smaller than the lower limit of CMIP5 resolution. Thus, the thermal effect of CO_2 forcing cannot be simulated, at least up to CMIP5 models and likely for the foreseeable future.

These concepts -- resolution, limits of accuracy, calibration uncertainty, and propagation of error -- are typically introduced in a freshman college year. And yet, the reviewer is evidently unfamiliar with each and all of them. Not to be unduly chagrinned though, as the author has yet to encounter a climate modeler conversant with these concepts.

3.2 Note that this term here is for climatology, not for change, as represented by cloud feedback. The author insist that his term "forcing" only represents radiative change during global warming against my reminder about the confusion of his usage, just because he did not correctly understand the term used in literature.

- 3.2.1 The author's references to cloud forcing encompass only the LWCF simulation uncertainty derived in [*Lauer and Hamilton*, 2013]. Nowhere in the manuscript, or in the author's first round response to this reviewer, are any calculations or representation made of changes in cloud forcing.
- 3.2.2 Note that the hindcasts, as used in [*Lauer and Hamilton*, 2013] to derive the LWCF calibration error, all represent climatology. Thus, the calibration error is representative of, and applicable to, climatological simulations.
- 3.2.3 The author does not, and did not, insist that forcing represents only radiative change.

Round 1 response 1, item 2.1 pointed out that manuscript line 144 (revised 146) described CO_2 radiative forcing; not change in forcing.

3.2.4 Manuscript eqn. 6 includes the change in CO₂ forcing with increasing [CO2]_{atm} due to human emissions. This identical change in CO₂ forcing drives the simulated change in air temperature presented in every single SRES and RCP projection scenario.

These are all standard usages in climate modeling. It is therefore difficult to understand the reviewer's confusion.

- 4.1 About model tuning. As I know, tuning in climate models is not to deal with errors in such a big magnitude, which is intolerable. All literatures the author cited is to deal with error in magnitude of ~10%, but not as huge as one that can be cumulated as ~100 times.
 - 4.1 The reviewer's concern is physical error, as in (simulated minus observed). However, there is no " ~100 times" physical error presented in the manuscript analysis.

The LWCF model calibration uncertainty is not physical error. Propagated LWCF calibration uncertainty does not signify physical error.

The large uncertainties show that the model projections do not transmit any predictive knowledge of future temperature.

This was all explained in crushing detail in manuscript lines 155-169, 551-567, 603-639, 681-693, 744-765, 800-822, 852-880, and extensively expanded in SI Section 7, *Differencing and Systematic Theory-bias Model Error* and SI Section 10, *The Meaning of Uncertainty in Model Projections*.

The reviewer apparently did not read any of this. Certainly, the reviewer has not displayed any knowledge of these analytical discussions.

This confusion of error with uncertainty ramifies through the reviewer's comments, and renders them without critical merit.

4.2 No one can tune a nonsense model to be close to observation, not even with data assimilation. The evidence to reject my point is invalid here.

- 4.2.1 The author has not concerned himself with nonsense models. The manuscript assesses accepted models. The model errors discussed in the manuscript are all known, and described in published literature.
- 4.2.2 At the center of reviewer comment 4 is the reviewer's persistent confusion of error with uncertainty. The propagated centennial uncertainty of ± 15 C is not physical error.

Indeed, the reviewer should have immediately understood that the ± 15 C cannot represent physical error because the air temperature 100 years from now is entirely unknown. No error can be calculated against an unknowable observable (air temperature 100 years from now).

Had the reviewer thought about the impossibility of calculating a simulation error against an unknowable observable, he would have realized his mistake in thinking this way. But, apparently, his thinking did not extend this far.

- 5.1 Again, the major problem of this paper is that the author misused a stable model error of ~10% as an error in trend, due to his artificially adding a unit of year-1.
 - 5.1 As noted in items 2.1 and 2.6, the year⁻¹ denominator derives directly from the method used in [*Lauer and Hamilton*, 2013] and stems directly from the standard method of root-mean-square statistical averages.

The year⁻¹ denominator is present in LWCF calibration uncertainty, and indeed is present by methodological necessity.

Response items 2.3 and 2.4 show that calculation of calibration uncertainties do not require a trend.

Nevertheless, the LWCF calibration uncertainty presented in [*Lauer and Hamilton*, 2013] do in fact derive from a 20-year trend in hindcast simulations.

- 5.2 His understanding of error propagation and accumulation starts from this wrong foundation, and goes nowhere.
 - 5.2 This reviewer's foundational premise is wrong. This reviewer's ability to distinguish between error and uncertainty is evidently absent. This reviewer shows no understanding of an empirical calibration experiment.

These mistakes misguide the entire review. The review therefore has no critical merit.

References:

Hubbard, K. G., and X. Lin (2002), Realtime data filtering models for air temperature measurements, Geophys. Res. Lett., 29(10), 1425 1421-1424; doi: 1410.1029/2001GL013191.

Lauer, A., and K. Hamilton (2013), Simulating Clouds with Global Climate Models: A Comparison of CMIP5 Results with CMIP3 and Satellite Data, J. Climate, 26(11), 3823-3845, doi: 10.1175/jcli-d-12-00451.1.

Pyle, J., et al. (2016), Chapter 1. Ozone and Climate: A Review of Interconnections, in Safeguarding the Ozone Layer and the Global Climate System: Issues Related to Hydrofluorocarbons and Perfluorocarbons, edited by T. G. Shepherd, S. Sicars, S. Solomon, G. J. M. Velders, D. P. Verdonik, R. T. Wickham, A. Woodcock, P. Wright and M. Yamabe, IPCC/TEAP Geneva.